

BAYESIAN NETWORKS IN PHILOSOPHY

STEPHAN HARTMANN, LUC BOVENS

1. INTRODUCTION

There is a long philosophical tradition of addressing questions in philosophy of science and epistemology by means of the tools of Bayesian probability theory (see Earman (1992) and Howson and Urbach (1993)). In the late '70s, an axiomatic approach to conditional independence was developed within a Bayesian framework. This approach in conjunction with developments in graph theory are the two pillars of the theory of Bayesian Networks, which is a theory of probabilistic reasoning in artificial intelligence. The theory has been very successful over the last two decades and has found a wide array of applications ranging from medical diagnosis to safety systems for hazardous industries.

Aside from some excellent work in the theory of causation (see Pearl (2000) and Spirtes et al. (2001)), philosophers have been sadly absent in reaping the fruits from these new developments in artificial intelligence. This is unfortunate, since there are some long-standing questions in philosophy of science and epistemology in which the route to progress has been blocked by a type of complexity that is precisely the type of complexity that Bayesian Networks are designed to deal with: questions in which there are multiple variables in play and the conditional independences between these variables can be clearly identified. Integrating Bayesian Networks into philosophical research leads to theoretical advances on long-standing questions in philosophy and has a potential for practical applications.

In the remainder of this contribution we will give a short introduction into the theory of Bayesian Networks (Sec. 2). We will then study one of the applications of Bayesian Networks in philosophy in more detail (Sec. 3) and finally discuss further possible applications and open problems (Sec. 4).

2. BAYESIAN NETWORKS IN ARTIFICIAL INTELLIGENCE

Bayesian Networks are a powerful tool to deal with probability distributions over a large class of variables if certain (conditional) independence relations between these variables are known. A probability

distribution over n variables contains 2^n entries. The number of entries will grow exponentially with the number of propositional variables. A Bayesian Network organizes these variables into a *Directed Acyclical Graph* (DAG), which encodes a range of (conditional) independences. A DAG is a set of nodes and a set of arrows between the nodes under the constraint that one does not run into a cycle by following the direction of the arrows. Each node represents a propositional variable. Consider a node at the tail of an arrow and a node at the head of an arrow. We say that the node at the tail is the parent node of the node at the head and that the node at the head is the child node of the node at the tail. There is a certain heuristic that governs the construction of the graph: there is an arrow between two nodes iff the variable in the parent node has a direct influence on the variable in the child node. From DAG to Bayesian Network, one more step is required. A Bayesian network contains a probability distribution for the variable in each root node (i.e. in each unparented node), and a probability distribution for the variable in each child node, conditional on any combination of values of the variables in their parent nodes. When implemented on a computer, a Bayesian network performs complex probabilistic calculations with one keystroke (see Cowell et al. (1999), Neapolitan (1990), and Pearl (1988)).

3. AN EXAMPLE: CONFIRMATION WITH AN UNRELIABLE INSTRUMENT

In philosophy of science, and more specifically in confirmation theory, there is a common idealization that the evidence in favor of a hypothesis is gathered by fully reliable instruments. What happens if we relax this idealization and permit that the evidence may have come from less than fully reliable instruments, as is common in scientific experimentation? Bayesian Networks prove to be useful to study situations like this. Consider a very simple scenario. Let there be a hypothesis, a (test) consequence of the hypothesis, a LTFR instrument and a report from the LTFR instrument to the effect that the consequence holds or not. To model this scenario, we need four propositional variables (written in *italic script*) and their values (written in roman script):

- (1) *HYP* can take on two values: HYP, i.e. the hypothesis is true and $\overline{\text{HYP}}$, i.e. the hypothesis is false;
- (2) *CON* can take on two values: CON, i.e. the consequence holds and $\overline{\text{CON}}$, i.e. the consequence does not hold;
- (3) *REL* can take on two values: REL, i.e. the instrument is reliable and $\overline{\text{REL}}$, i.e. the instrument is not reliable;

- (4) *REP* can take on two values: *REP*, i.e. there is a positive report, or, in other words, a report to the effect that the consequence holds and \overline{REP} , i.e. there is a negative report, or, in other words, a report to the effect that the consequence does not hold.

A probability distribution over these variables contains 24 entries. To represent the information in a more parsimonious format, we construct a Bayesian Network. Following our heuristics, we can construct a Bayesian Network for the case at hand. Whether the consequence holds is directly influenced by and only by whether the hypothesis is true or not; whether there is a report to the effect that the consequence holds is directly influenced by and only by whether the consequence holds or not and by whether the instrument is reliable or not. Hence, we construct the basic graph in figure 1 in which the node with the variable *HYP* is a parent node to the node with the variable *CON* and the nodes with the variables *CON* and *REL* are in turn parent nodes to the node with the variable *REP*. We stipulate prior probability distributions for the variables in the root nodes of the graph

$$(1) \quad P(HYP) = h, \quad P(REL) = r,$$

with $0 < h, r < 1$, and conditional probability distributions for the variables in the other nodes given any combination of values of the variables in their respective parent nodes. Consider the node with the variable *CON* which is a child node to the node with the variable *HYP*. We take a broad view of what constitutes a consequence, that is, we do not require that the truth of the hypothesis is either a necessary or a sufficient condition for the truth of the consequence. Rather, a consequence is to be understood as follows: the probability of the consequence given that the hypothesis is true is greater than the probability of the consequence given that the hypothesis is false:

$$(2) \quad P(CON|HYP) = p > q = P(CON|\overline{HYP}).$$

Consider the node with the variable *REP*, which is a child node to the nodes with the variables *CON* and *REL*. How can we model the workings of an unreliable instrument? Let us make an idealization: We suppose that we do not know whether the instrument is reliable or not, but if it is reliable, then it is fully reliable and if it is not reliable, then it is fully unreliable. Let a fully reliable instrument be an instrument that provides maximal information: it is an instrument that says of

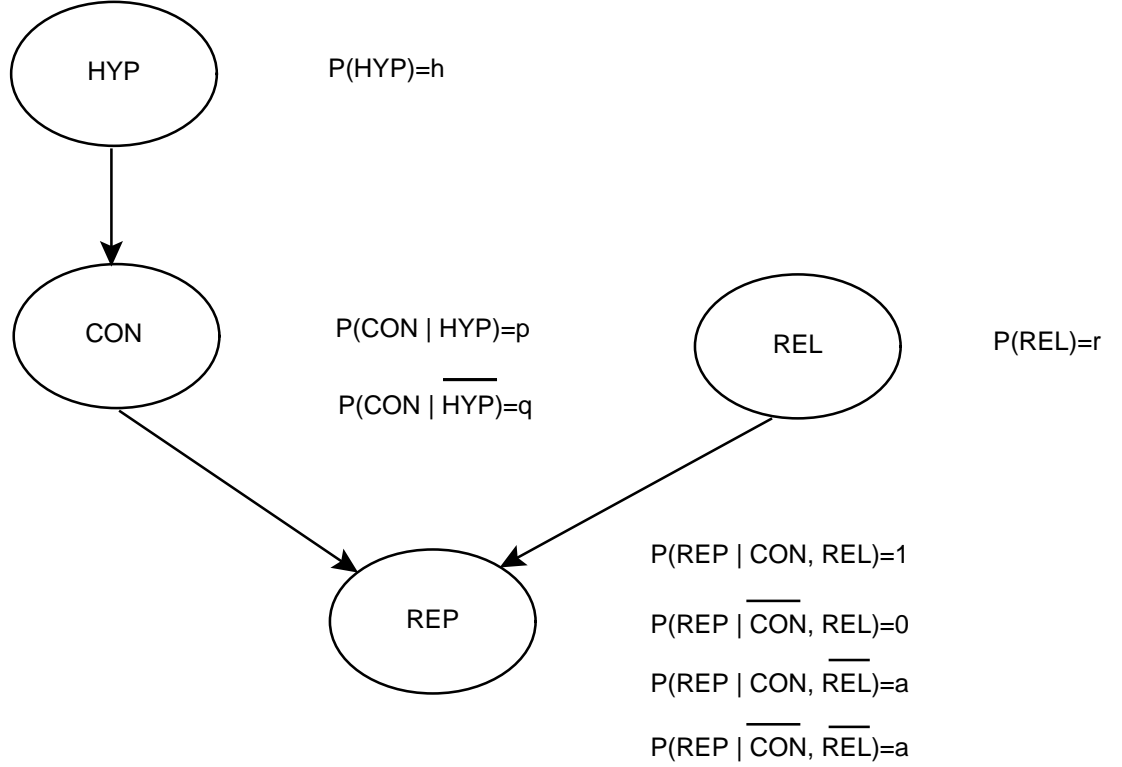


FIGURE 1. Bayesian Network for the confirmation of a hypothesis with an unreliable instrument.

what is that it is, and of what is not that it is not:

$$(3) \quad P(REP | REL, CON) = 1 \quad \text{and} \quad P(REP | REL, \overline{CON}) = 0.$$

Let a fully unreliable instrument be an instrument that provides minimal information: it is an instrument that is no better than a randomizer:

$$(4) \quad P(REP | \overline{REL}, CON) = P(REP | \overline{REL}, \overline{CON}) = a$$

with $0 < a < 1$ where a is called the randomization parameter.

We can now construct the Bayesian Network by adding the probability values to the graph in figure 1. What's so great about Bayesian Networks? A central theorem in the theory of Bayesian Networks states that a joint probability distribution over any combination of values of the variables in the Network is equal to the product of the probabilities and conditional probabilities for these values as expressed in the Network. For example, suppose we are interested in the joint probability of HYP, \overline{CON}, REP and \overline{REL} . We can read the joint probability

directly off of figure 1:

$$\begin{aligned} P(\text{HYP}, \overline{\text{CON}}, \text{REP}, \overline{\text{REL}}) &= P(\text{HYP})P(\overline{\text{REL}})P(\overline{\text{CON}}|\text{HYP})P(\text{REP}|\overline{\text{CON}}, \overline{\text{REL}}) \\ &= h(1-r)(1-p)a \end{aligned}$$

Standard probability calculus teaches us how to construct marginal distributions out of joint distributions and subsequently conditional distributions out of marginal distributions. When implemented on a computer, Bayesian Networks provide a direct answer to such queries. We are interested in the probability of the hypothesis given that there is a report from a LTFR instrument that the consequence holds. This probability is

$$(5) \quad P^*(\text{HYP}) = P(\text{HYP}|\text{REP}) = \frac{P(\text{HYP}, \text{REP})}{P(\text{REP})}.$$

For ease of representation, we will abbreviate $\bar{x} := 1 - x$ for all variables x .

$$(6) \quad P^*(\text{HYP}) := P(\text{HYP}|\text{REP}) = \frac{h(pr + a\bar{r})}{hr(p - q) + qr + a\bar{r}}$$

We measure the degree of confirmation that the hypothesis receives from a positive report by the difference:

$$(7) \quad C(\text{H}) := P^*(\text{HYP}) - P(\text{HYP}) = \frac{h\bar{h}(p - q)r}{hr(p - q) + qr + a\bar{r}}$$

4. FURTHER APPLICATIONS AND OPEN PROBLEMS

We know now how to model the degree of confirmation that a hypothesis receives from a single positive report concerning a single consequence of the hypothesis by means of a single LTFR instrument. This basic model can function as a paradigm to model complex strategies to improve the degree of confirmation that can be obtained from LTFR instruments. We can raise the following questions: What is the impact of repeating the experiment many times over? Of repeating the experiment with different instruments? Of developing a theoretical underpinning that boosts the reliability of the instrument? Of calibrating the instrument? These are the sort of questions that can be fruitfully modeled by means of Bayesian Networks. Careful mathematical analysis of these models yields surprising results about, say, the variety-of-evidence thesis, the Duhem-Quine thesis and about calibration procedures. For details see Bovens and Hartmann (2001a). Future problems to be addressed in this direction concern the inclusion of an arbitrary number of consequences and reports, as well as a more realistic modeling of the reliability of test instruments. All this

can be done in the context of Bayesian Networks in an intuitive and computationally powerful way.

Bayesian Networks also have a variety of applications in other parts of philosophy. In epistemology, for example, foundational questions have not been addressed sufficiently within a probabilistic framework. There is the sceptical challenge dating back to Descartes? Meditations that we cannot trust our senses and that our empirical knowledge has no justification. The coherentist answer is that even though the processes by means of which we gather information about the world may be less than fully reliable, the very fact that the scientific story fits together, i.e. has an internal coherence, provides justification that the story is true. But how are we to understand the claim that our information gathering processes are less than fully reliable? How are we to understand the claim that a story is internally coherent? There are many open questions about these central notions in coherentism. Within the framework of Bayesian Networks, multiple notions of less-than-full reliability can be modeled (see Bovens and Olsson (2000)) and a probabilistic measure of coherence, which has been a long-time dream of coherentists, can be developed (see Bovens and Hartmann (2001b)). With a clear understanding of these central notions in hand, the coherentist answer to the Cartesian sceptic can be assessed.

This theoretical work on reliability and coherence has practical applications in the theory of belief change. How does a cognitive system (a person or an expert system) update its beliefs when it receives new information as its input? Under what conditions does it add this new item of information to its previous beliefs? Under what conditions does it discard some of its previous beliefs? On the standard approach, new information is added to the belief set until an inconsistency appears. It is more realistic to let belief change be determined by two factors, viz. how reliable are our information sources and how well does the new information cohere with what we already believe. These factors can be directly modeled by Bayesian Networks. Such modeling yields novel theoretical insights about belief change and carries a promise of applications to information management in expert systems (for details see Bovens and Hartmann (2000)).

REFERENCES

- Bovens and Hartmann (2000).**: Luc Bovens and Stephan Hartmann, "Coherence, Belief Expansion and Bayesian Networks" Proceedings of the 8th International Workshop on Non-Monotonic

- Reasoning, NMR'2000, Breckenridge, Colorado, USA, April 9-11, 2000. (<http://www.cs.engr.uky.edu/nmr2000/proceedings.html>)
- Bovens and Hartmann (2001a).**: Luc Bovens and Stephan Hartmann, "Bayesian Networks and the Problem of Unreliable Instruments." To appear in *Philosophy of Science*. (<http://philsci-archive.pitt.edu/documents/disk0/00/00/00/95/index.html>)
- Bovens and Hartmann (2001b).**: Luc Bovens and Stephan Hartmann, "Solving the Riddle of Coherence." Unpublished manuscript.
- Bovens and Olsson (2000).**: Luc Bovens and Erik J. Olsson, "Coherentism, Reliability and Bayesian Networks" In: *Mind* 109, Issue 436, pp. 685-719.
- Cowell et al. (1999).**: Robert Cowell, Philip Dawid, Steffen Lauritzen, and David Spiegelhalter, *Probabilistic Networks and Expert Systems*. New York: Springer.
- Earman (1992).**: John Earman, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: MIT Press.
- Howson and Urbach (1993).**: Colin Howson and Peter Urbach, *Scientific Reasoning - The Bayesian Approach*. (2nd ed.) Chicago: Open Court.
- Neapolitan (1990).**: Richard Neapolitan, *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. New York: Wiley.
- Pearl (1988).**: Judea Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Cal.: Morgan Kaufmann.
- Pearl (2000).**: Judea Pearl, *Causation: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Spirtes et al. (2001).**: Peter Spirtes, Clark Glymour, and Richard Scheines *Causation, Prediction, and Search*. Cambridge, Mass.: MIT Press.

(S. Hartmann) CENTER FOR PHILOSOPHY OF SCIENCE, UNIVERSITY OF PITTSBURGH, 817 CATHEDRAL OF LEARNING, PITTSBURGH, PA 15260, USA
E-mail address, S. Hartmann: shart@pitt.edu

(L. Bovens) UNIVERSITY OF COLORADO AT BOULDER, DEPARTMENT OF PHILOSOPHY, CB 232, BOULDER, CO 80309, USA
E-mail address: bovens@spot.colorado.edu